Some Remarks on the Distribution of twin Primes

Marek Wolf

Institute of Theoretical Physics, University of Wrocław
Pl.Maxa Borna 9, PL-50-204 Wrocław, Poland
e-mail: mwolf@ift.uni.wroc.pl

Abstract

The computer data up to $2^{44} \approx 1.76 \times 10^{13}$ on the gaps between consecutive twins is presented. The simple derivation of the heuristic formula describing computer results contained in the recent papers by P.F.Kelly and T.Pilling [5], [6] is provided and compared with the "experimental" values.

Key words: *Prime numbers,twins*
MSC: 11A41 (Primary), 11Y11 (Secondary)

Among the primes the subset of twin primes is distinguished: twins are such numbers (p, p') that both p and $p' = p+2$ are prime. So the set of twins starts with (3, 5), (5, 7), (11, 13), (17, 19), (29, 31) .... It is not known whether there is infinity of twins; the largest known today pair of twins was found recently by Underbakke, Carmody, Gallot (see http://www.utm.edu/research/primes/largest.html) and it is the following pair of 29603 digits numbers:

$$1807318575 \times 2^{98305} \pm 1. \tag{1}$$

The mathematicians are using the notation $\pi_2(N)$ to denote the number of twins smaller than N and the Hardy and Littlewood conjecture B states [2] that the number of twins below a given bound N should be approximately equal to

$$\pi_d(N) \sim c_2 \int_2^N \frac{du}{\ln^2(u)} = c_2 \frac{N}{\ln^2(N)} + \dots . \tag{2}$$

where the constant c_2 (sometimes called "twin–prime" constant) is defined in the following way:

$$c_2 \equiv 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) = 1.32032\dots \tag{3}$$

The problem of distribution of twins is very difficult. One of the characteristics one can use for this purpose is the statistics of gaps between consecutive twins. Let d denote the distance between two consecutive twins measured as a arithmetical difference between last primes constituting consecutive twins, thus for example for twins (29,31) and (17,19) $d = 12$. The distances can be only multiplicities of 6: $d = 6k$, because all twins are of the form $6k \pm 1$[4]. Next let $m(d, N)$ denote the number of twins separated by d and smaller than N:

$$m(d, N) = \text{number of consecutive twins } (p_m, p_{m+1} = p_m + 2) \text{ and } (p_n, p_{n+1} = p_n + 2)$$

$$\text{such that } p_n - p_{m+1} = d, \ \textit{and } p_{n+1} < N \tag{4}$$

A few years ago I have obtained on the computer data for $m(d, N)$ for $N = 2^{26}, \dots N = 2^{44}$ [3]. A part of results is shown on the Fig.1.

Recently there appeared two papers by P.F.Kelly and T.Pilling [5], [6]. They have looked on the histogram of separations between consecutive twins measured by the *number of primes* in between, not just the arithmetical difference as above. Here I am providing heuristic formulas which describe findings of Kelly and Pilling as well as much larger than in [5] and [6] computer data which corroborates the obtained analytical relations.

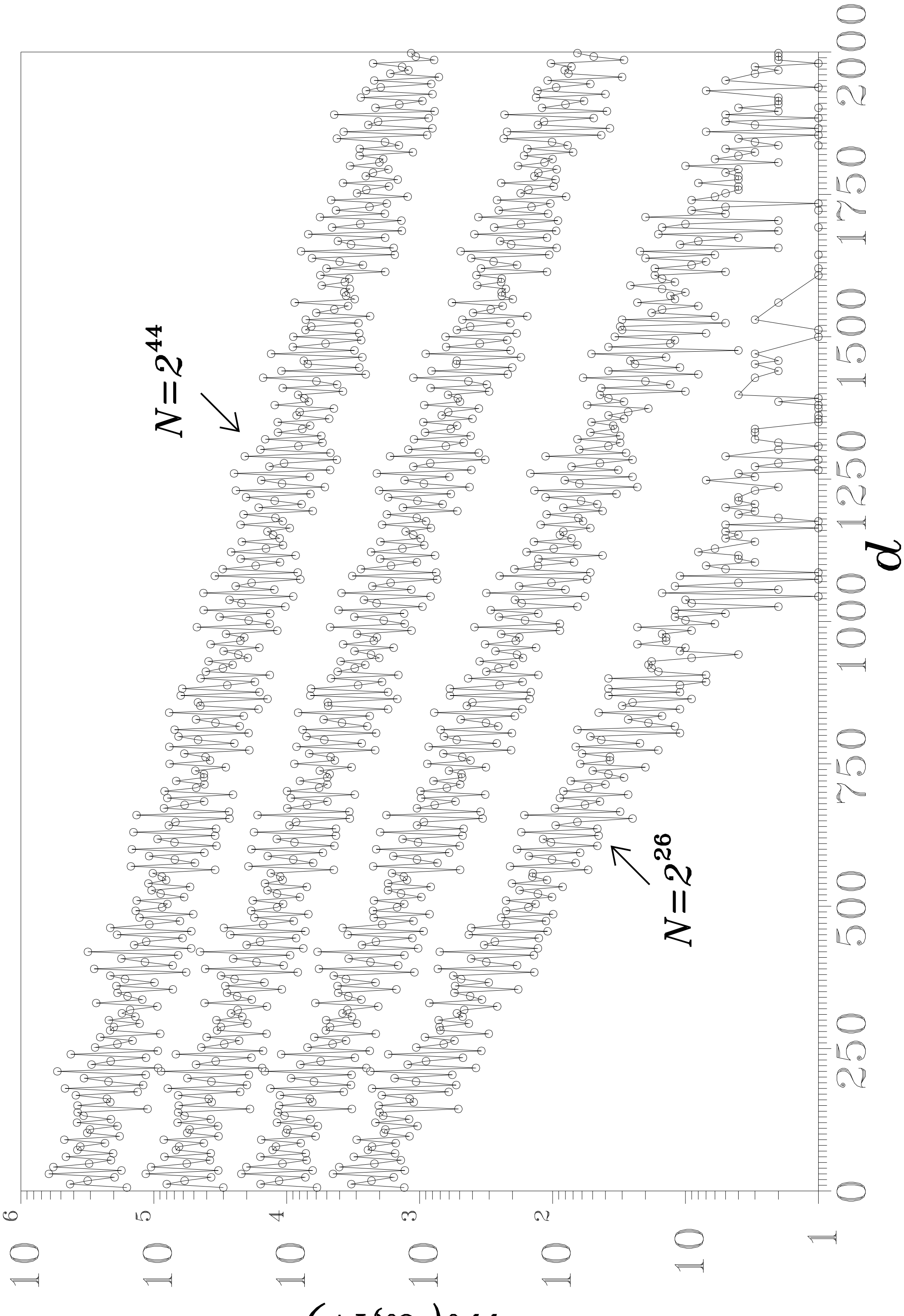

Figure 1: The plot showing the dependence of histogram of gaps between consecutive twins at $N = 2^{26}, 2^{32}, 2^{38}$ and 2^{44}. There is linear scale on x axis and logarithmical scale on y scale.

Let $\mu(s,N)$ denote the number of consecutive twins smaller than N and separated by s primes, with exception of primes constituting twins in question, i.e. for example for (5, 7) and (11, 13) $s=0$ while for (17, 19) and (29, 31) $s=1$ (prime 23 lies in between). More strictly we adopt the following definition:

$$\mu(s,N) \;=\; number\ of\ consecutive\ twins\ (p_m, p_{m+1}=p_m+2)\ and\ (p_n, p_{n+1}=p_n+2)$$
$$such\ that\ \pi(p_n)-\pi(p_{m+1})-1=s,\ \ and\ p_{n+1}<N \tag{5}$$

Here $\pi(N)$ is as usual the number of primes$<N$ and

$$\pi(N)=\int_2^N \frac{du}{\ln(u)}=\frac{N}{\ln(N)}+\ldots. \tag{6}$$

The authors of [5] have found that $\mu(s,N)$ for a given N decreases exponentially with s. In fact these authors are working with relative frequencies $\mu(s,N)/\sum_t \mu(t,N)$ but here I will use absolute values $\mu(s,N)$.

I have made the computer search up to $N=2^{44}\approx 1.76\times 10^{13}$ and counted the number of primes between consecutive twins. During the computer search the data representing the function $\mu(s,N)$ were stored at values of N forming the geometrical progression with the ratio 4, i.e. at $2^{22},\ldots,2^{42},2^{44}$. The resulting curves are plotted in the Fig.2.

Because the points lie on the straight lines on the semi-logarithmic scale, we can infer from the Fig.2 the ansatz

$$\mu(s,N)\sim A(N)e^{-B(N)s}. \tag{7}$$

The functions $A(N)$ and $B(N)$, giving the intercepts and the slopes of straight lines seen in the Fig.2, can be determined by exploiting two identities that $\mu(s,N)$ have to obey. First of all, the sum of $\mu(s,N)$ over all s is the total number of twins $<N$:

$$\sum_{s=0}^{s_{max}(N)} \mu(s,N) \;=\; \pi_2(N). \tag{8}$$

Here $s_{max}(N)$ is the largest separation s between twins $<N$, see later discussion. The second selfconsistency condition comes from the observation, that

$$\sum_{s=0}^{s_{max}(N)} s\mu(s,N)=\pi(N)-2\pi_2(N). \tag{9}$$

In fact the above sum starts with $s=1$ what is important for the relations obtained below. Putting the ansatz (7) into (8) and (9) and collecting appropriately terms we end up with the geometrical series with quotient $e^{-B(N)}$ and (9) is a differentiated

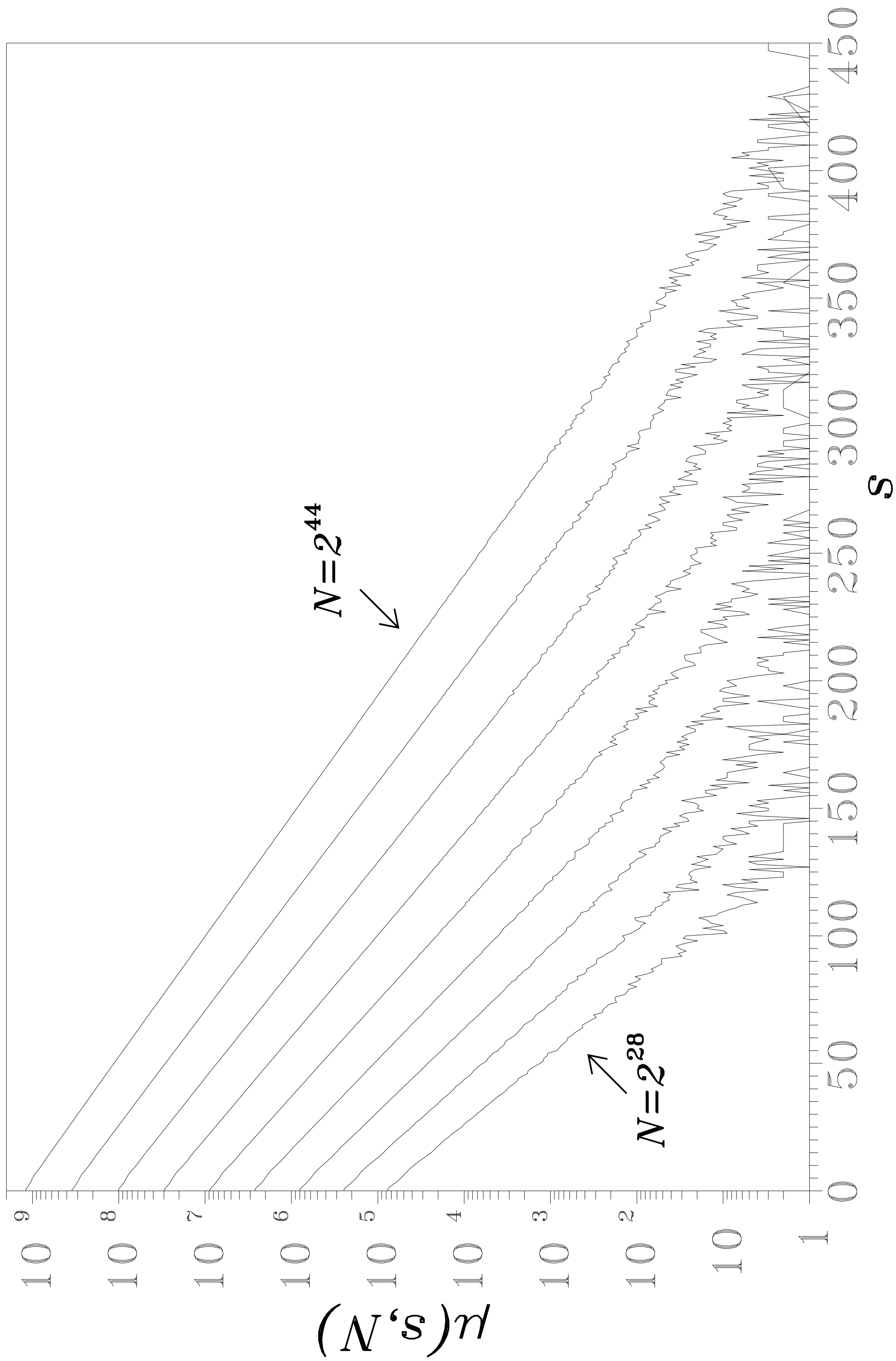

Figure 2: The plot showing the dependence of the histogram $\mu(s, N)$ on s at $N = 2^{28}, 2^{30}, \ldots, 2^{44}$. There is a logarithmical scale on the y-axis , while on the x-axis there is a linear scale.

geometrical series. Using the approximation and summing in (8) and (9) up to infinity instead of particular s_{max} (which depends on N) I end up with the equations:

$$\frac{A(N)}{1-\exp\{-B(N)\}} = \pi_2(N) \tag{10}$$

and

$$\frac{A(N)e^{-B(N)}}{(1-\exp\{-B(N)\})^2} = \pi(N) - 2\pi_2(N). \tag{11}$$

From the Fig.2 it is seen that the slopes $B(N)$ are smaller than 1 and they decrease with N, e.g. $B(2^{42}) = 0.0503...$, $B(2^{44}) = 0.0483....$ Thus in the above relations I can put $\exp\{-B(N)\} \approx 1 - B(N)$, $B(N) \ll 1$ and solve the equations. Finally the main conjecture is:

$$A(N) \approx \frac{\pi_2^2(N)}{\pi(N) - 2\pi_2(N)}, \qquad B(N) \approx \frac{\pi_2(N)}{\pi(N) - 2\pi_2(N)}. \tag{12}$$

Putting here asymptotic forms of $\pi(N)$ and $\pi_2(N)$ I make the following guess:

$$A(N) \sim \frac{c_2^2 N^2}{\ln^3(N)}, \qquad B(N) \sim \frac{c_2}{\ln(N)}. \tag{13}$$

The comparison of these formulae is given in Table 1, where the values $A_{theor}(N)$ and $B_{theor}(N)$ obtained from (10) and (11) are divided by values $A_{exp}(N)$ and $B_{exp}(N)$ obtained by least-square method from data presented on Fig.2. The actual numbers A_{exp} and B_{exp} depend on the number of points taken for linear regression method, results in Table 1 are obtained when I skipped 15 first points and 40% of last point, where large fluctuation appear (in fact only 2 points determine straight line). For the calculation of $A_{theor}(N)$ and $B_{theor}(N)$ I have used *exact* values of $\pi(N)$ and $\pi_2(N)$ at $N = 2^{22}, \ldots N = 2^{44}$ from my earlier computer run [8].

All ratios tend to 1 with increasing N, as it should be and asymptotic values (13) are smaller than exact values, since in (2) and (6) dots denote positive terms which were skipped.

Another characteristic which can be used to test the conjecture (12) is the question what are the largest separations $s_{max}(N)$ between twins up to a given N. It corresponds to problem of the maximal gaps between consecutive primes, what has a long history, see [7]. Heuristic argument is that the maximal gap appears only once and hence it can be obtained from the equation:

$$\mu(s_{max}(N), N) = 1 \tag{14}$$

TABLE **I**

N	$A_{exp}(N)/A_{theor}(N)$	$B_{exp}(N)/B_{theor}(N)$	$A_{exp}(N)/A_{asympt}(N)$	$B_{exp}(N)/B_{asympt}(N)$
2^{22}	0.964932	0.971223	1.528465	1.308701
2^{24}	0.983914	0.986198	1.475196	1.283024
2^{26}	0.967421	0.980584	1.389283	1.241010
2^{28}	0.972147	0.982905	1.353625	1.219132
2^{30}	0.965933	0.980277	1.312914	1.196643
2^{32}	0.970505	0.983126	1.291164	1.183385
2^{34}	0.973114	0.984629	1.270698	1.170806
2^{36}	0.975160	0.985884	1.252173	1.159566
2^{38}	0.976340	0.986722	1.235622	1.149610
2^{40}	0.977286	0.987373	1.220839	1.140667
2^{42}	0.977457	0.987681	1.206974	1.132435
2^{44}	0.978554	0.988279	1.195782	1.125445

From the main conjecture (12) it follows that for large N:

$$s_{max}(N) \approx \frac{\pi(N)}{\pi_2(N)}\Big(2\ln\pi_2(N) - \ln(\pi(N) - 2c_2\pi_2(N))\Big) \tag{15}$$

For large N it goes into the

$$s_{max}(N) \sim \frac{1}{c_2}\ln^2(N) \tag{16}$$

what differs from the Cramer's conjecture for maximal gaps between consecutive primes by the factor $1/c_2$. It differs also by on power of $\ln(N)$ from the asymptotic behavior of maximal gaps between consecutive twins $G_2(N) \sim \ln^3(N)$ obtained just arithmetically: i.e. distance between twins $p_m, p_{m+1} = p_m + 2$ and $p_n, p_{n+1} = p_n + 2$ is simply $p_n - p_m$ and not $\pi(p_n) - \pi(p_{m+1}) - 1$. The comparison of actual values of $s_{max}(N)$ obtained from the computer search are compared with formula (15) in the Fig.3.

Finally I will make a remark on the problem of champions, i.e. the most often occurring gaps. For prime numbers it was treated in [9]. In [6] the authors have made the assertion that the champions for twins with gaps measured in terms of the primes lying in between is always $s = 0$. From the Fig.1 it is seen that when gaps between twins are arithmetical differences then there will be a set of champions consisting of $d = 30$, next emerging peak is $d = 210$ and so on, as it can be seen on the data presented in Fig.1. It will be discussed in more detail in the forthcoming paper.

Finally let us notice that from the Fig. 2 it follows that the separations between twins measured by the number of primes in between follow exactly the Poissonian

behavior, see e.g. [10]. Interestingly the change in the “measuring sticks” removes oscillations from Fig.1 and leaves pure exponential decrease.

References

[1] P.Billingsley, "Prime Numbers and Brownian Motion", *Amer. Math. Monthly* **80** (1973), pp.1099-1115

[2] G.H.Hardy and J.E. Littlewood, *Acta Mathematica* **44** (1922), p.1-70

[3] M.Wolf, 1999, unpublished

[4] Cuesta-Dutari,-Norberto *Arithmetic of the sequences* $6n-1$, $6n+1$ *and of twin primes.* Collect.-Math. [Consejo-Superior-de-Investigaciones-Cientificas.- Universidad-de-Barcelona.-Collectanea-Mathematica.-Seminario-Matematico-de- Barcelona] 37 (1986), no. 3, p.211–227

[5] P.F. Kelly and T.Pilling, *Characterization of the Distribution of twin Primes*, arXiv:math.NT/0103191

[6] P.F. Kelly and T.Pilling, *Implication of a New Characterization of the Distribution of twin Primes*, arXiv:math.NT/0104205

[7] H.Cramer, "On the order of magnitude of difference between consecutive prime numbers", *Acta Arith.* **2** (1937), p.23-46; D.Shanks, "On Maximal Gaps between Succesive Primes", *Math.Comp.* **18** (1964), p.464; J.H.Caldwell, "Large Intervals Between Consecutive Primes", *Math.Comp.***25** (1971), p.909; L.J.Lander and T.R.Parkin, "On First Apperance of Prime Differences", *Math.Comp.* **21** (1967), p.483; R.P.Brent, "The First Occurrence of Certain Large Prime Gaps", *Math.Comp.***35** (1980), p.1435-1436; J.Young and A.Potler, "First occurence Prime Gaps", *Math.Comp.* **52** (1989), p.221-224;

[8] M.Wolf, *Some conjectures on the gaps between consecutive primes*, preprint IFTUWr 894/95, available at http://www.ift.uni.wroc.pl/∼mwolf

[9] A. Odlyzke, M. Rubinstein and M. Wolf, *Jumping Champions*, Exp. Math., **8** (1999) 107-118. Also available at http://www.ift.uni.wroc.pl/∼mwolf and http://www.research.att.com/∼amo/doc/numbertheory.html

[10] O. Bohigas and M.J. Giannoni, in *Mathematical and Computational Methods in Nuclear Physics*, edts. J.S. Dehesa et. al. (Springer, Heidelberg, New York, 1984)

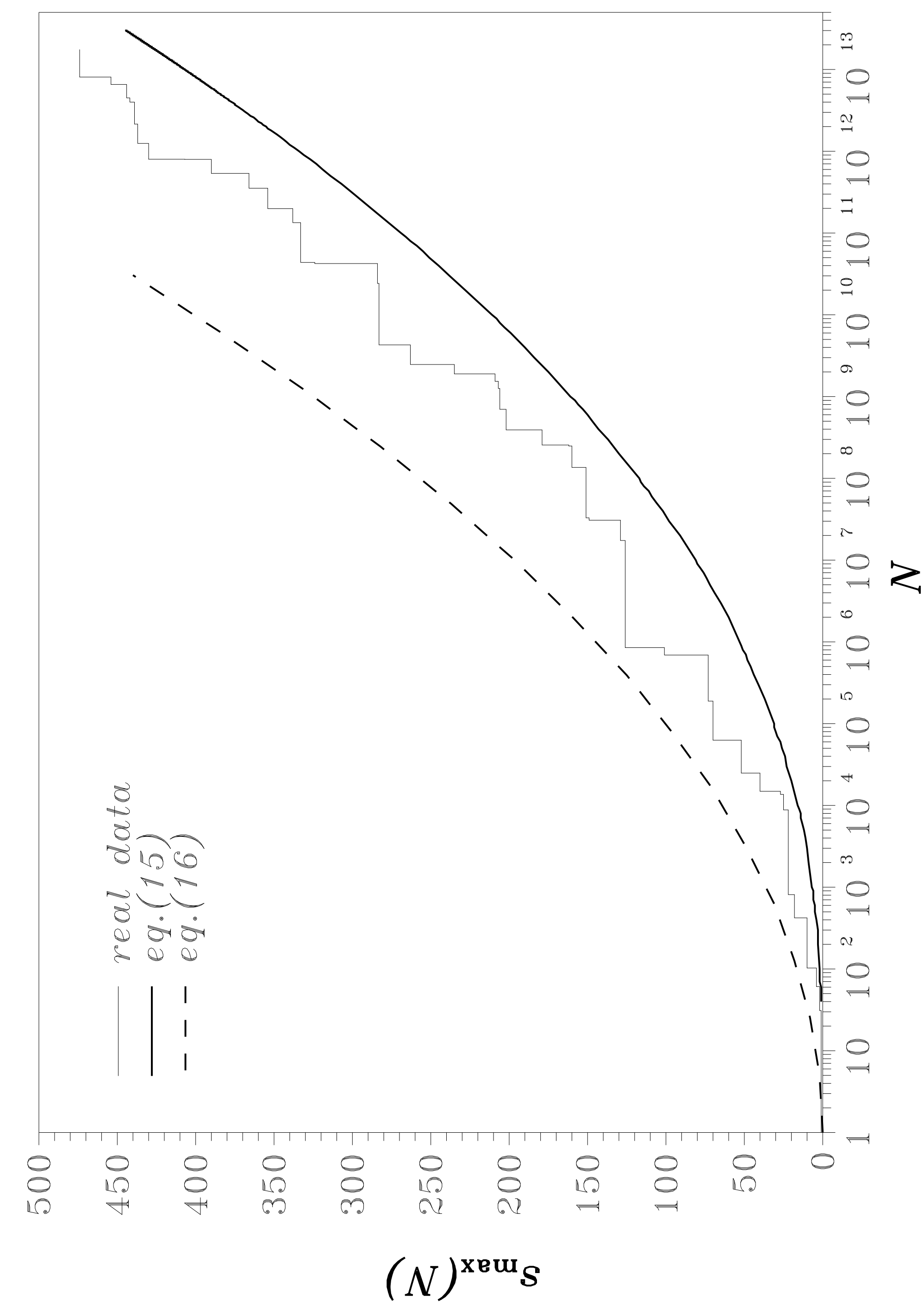

Figure 3: The plot showing the dependence of $s_{max}(N)$ and comparison with formula (15).